Lucia Ku
Professor Avinash Jairam
CIS 3120
12 March 2022

Homework 1 Report

For this assignment, I used Google Colaboratory to create a code using Python to find the average heights of each listed teams' players. In the first coding box, I first imported the Beautiful Soup library, which will be used to parse and extract the data from each HTML code of each website. Then, I imported the requests library, which is commonly used to help web scrape data in combination with Beautiful Soup. Lastly, I imported the statistics library in order to find the averages of the heights scraped from each website at the end of the code.

In the second coding box, I created a dictionary called "sports_teams" and proceeded to list each url inside of the brackets. I gave each key in the dictionary a name that corresponds with the information provided in the websites, and listed each website's URL as the value of the dictionary. Additionally, for this coding box, I used the "An Efficient Approach to HW1" that was provided to us as a reference.

In the third coding box, I first defined a new function called "scrape" in order to scrape the websites in the previous boxes. Next, I made a request to the server to get the URL key from the dictionary and put this request into "page" to simplify it. Then, I imported the raw HTML into Beautiful Soup using the "page" from the previous line and adding a ".content" to fetch the contents of each URL. I also put that into "soup" to simplify it. Next, I visited one of the websites, right clicked, and viewed its page source in order to find out what tag and class was used to list the heights on the website. In order to do this, I pressed "ctrl" and "f" together and searched for "height." After finding out the tag and class used, I used the "soup" from the previous line and added a ".find_all" in order to find every td tag with the class of "height" on the websites. I also put that in "all_relevant_tags" to simplify it. Next, I created an empty list with the name of "average" in order to store all the averages in. Then, I created a "for loop" to loop through the dictionary in the previous box so as to make it simpler to go through each website without having to repeatedly rewrite the same codes for each one. I specifically looped the "height" in "all_relevant_tags." Underneath the for loop, I used "height.get_text().split('-')" to first get the heights of each player and then convert them into a list while splitting the feet from the inches using "-" to define how to split the height, since the heights in "height.get_text()" alone has split the feet and inches with the "-" mark. (I used the "Guide to Homework 1" as a reference for this line). Next, I used "average.append(int(h[0])*12+int(h[1]))" to convert the string of the feet measurement into a number and multiplied it by 12 in order to convert feet to inches. Then, I converted the string of the inches measurement into a number and added that to the previously converted feet measurement for a total amount of height in inches. Lastly, I put in a command to print the averages of each sports team with the statistics library and calculated it using the "statistics.mean(average)". I also used the "An Efficient Approach to HW1" as reference for this box and all of this information was indented underneath the first "def scrape" line.

In the last coding box, I used a for loop to run through each website in the dictionary in the second box by using "sports_teams.items()" which gets all the "key, value" pairs in the dictionary. Then, underneath the for loop, I call the scrape function that I just defined in the previous box to call all the website pairs in the dictionary.

Lastly, here is a summary of the results and some interpretations of the data below based on the questions listed in the Homework #1 pdf file:

1) The average height of the men's swimming team is 71.53 inches.
2) The average height of the men's volleyball team is 73.27 inches.
3) The average height of the women's swimming team is 64 inches.
4) The average height of the women's volleyball team is 65.33 inches.
5) The average height of the men's volleyball team is around 2 inches taller than the average height of the men's swimming team. Overall, there does not seem to be a drastic difference in height between both teams.
6) The average height of the women's volleyball team is around 1 inch taller than the average height of the women's swimming team. Overall, there does not seem to be a drastic difference in height between both teams.
7) According to this exercise and the data gathered specifically from Baruch's swimming and volleyball teams, the average swimmer actually seems to be slightly shorter than the average volleyball player. However, overall, there does not seem to be a drastic difference in height between both sports.